

# Ferramenta para rotulação de séries temporais geradas por dados de acelerômetros

Juan Carlos Santos Silva<sup>1</sup>, Rafael Geraldelli Rossi<sup>1</sup>, Vinícius Mourão Alves de Souza<sup>2</sup>

<sup>1</sup>Universidade Federal de Mato Grosso do Sul (UFMS)  
Caixa Postal 210 – 79613-000 – Três Lagoas – MS – Brasil

<sup>2</sup>Onion Tecnologia  
13560-460 – São Carlos – SP – Brasil

jjcarlos853@gmail.com, rafael.g.rossi@ufms.br, vinicius.email@gmail.com

**Resumo.** Nos dias atuais há um grande volume de tráfego terrestre, sendo o maior responsável pelo transporte de cargas e pessoas no país. Monitorar e manter essas vias é um grande desafio para os órgãos responsáveis. Na maioria das vezes, a inspeção manual é preferível para realizar o monitoramento das vias, pois acredita-se ser o modo mais econômico financeiramente ou mesmo desconhece-se outras formas de monitoramento. Porém, nos dias atuais existem equipamentos eletrônicos, operados manualmente, os quais facilitam o processo de monitoramento. No entanto, estes equipamentos possuem alto custo financeiro e de complexidade de operação, além de restrições de distância de monitoramento. Para suprir este problema, foi proposta uma aplicação para smartphones onde se utiliza o acelerômetro dos mesmo para captar, em tempo real, possíveis problemas relacionados a qualidade do asfalto. O acelerômetro irá coletar dados como o tempo, a posição no sistema de posicionamento global, a magnitude, a latitude, a longitude, a velocidade e o deslocamento, que são capturados por ele e utilizados neste trabalho, onde o campo de magnitude será plotado em um gráfico na ordem dada pelo campo de tempo. Esses dados serão submetidos à um classificador o qual deverá classificar, por exemplo, "Trecho normal" ou "Trecho com buraco". Uma forma viável de se gerar um classificador é por meio do aprendizado de máquina. Porém, para gerar tal rotulador, é necessário que se apresentem dados rotulados com as categorias do problema a ser tratado, para que o algoritmo de aprendizado de máquina possa extrair padrões das classes e gerar um modelo de classificação, o qual irá classificar os dados oriundos do acelerômetro. Na literatura existem aplicações que realizam a rotulação de séries temporais, no entanto essas aplicações não permitem sincronizar os dados com multimídias que facilitem o entendimento dos dados. Dado isso, foi proposto neste trabalho de conclusão de curso uma ferramenta que permite ao usuário rotular os dados oriundos do acelerômetro, podendo visualizar as vias sincronizadamente por meio de vídeos fornecidos pelo usuário.

## 1. Introdução

Grandes investimentos e uma política estatal adotada a partir da década de 50 levaram a malha rodoviária à preferência no transporte de pessoas e bens no Brasil. De acordo com a estimativa do Ministério dos Transportes [Brasil 2014], a malha rodoviária é responsável por 61% das movimentações do setor de transporte de cargas do país, sendo esses valores

ainda mais significativos no setor de transporte de passageiros, abrangendo 95% de participação. Essa opção de locomoção se tornou muito cara devido as manutenções gerados pela má qualidade das rodovias, afetando diretamente a economia do país, ao passo que o custo dos transportes é um dos maiores agregadores de preço aos bens. Na malha asfáltica urbana, o impacto é ainda maior devido ao uso diário da população em seu cotidiano, uma vez que ela é utilizada para se chegar ao trabalho, ir a escola, fazer compras e até mesmo o lazer. Apesar de estar em queda, a frota circulante de veículos leves nas cidades atinge mais de 35 milhões, levando a malha asfáltica urbana a ser a via mais importante de locomoção da população urbana brasileira [Bazani 2017].

Mesmo considerando a importância da malha asfáltica, a porcentagem de estradas pavimentadas no Brasil atinge pouco mais de 12% [Brasil 2017]. Apesar de pequena extensão, a qualidade da mesma também é questionada, tendo que a qualidade das rodovias brasileiras ocupa a 121ª posição de um total de 140 países analisados no *ranking* de competitividade global do Fórum Econômico Mundial, divulgado em setembro de 2015 [CNT 2014a]. Segundo a Pesquisa de Rodovias da Confederação Nacional do Transporte (CNT), 62% das rodovias possuem um estado péssimo, ruim ou regular. Ainda de acordo com a CNT, o custo de operações dos veículos em estradas de má qualidade aumentam cerca de 25% em relação às rodovias de ótima qualidade [CNT 2014a]. Além disso, o Brasil tem prejuízo anual de 3,8 bilhões de reais somente com o transporte de grãos de soja para exportação devido a problemas nas rodovias [CNT 2014b]. Vale ressaltar que além do mal-estar econômico, a qualidade das vias também afetam a segurança dos motoristas e pedestres.

Diversos motivos colaboram para o deterioramento do asfalto, como grande fluxo de veículos, excesso de chuvas e má qualidade de pavimentação, fazendo com que as vias sofram exaustivas perdas de qualidade. Existem órgãos públicos e privados responsáveis pela manutenção e melhoria do pavimento das vias. Uma das formas utilizadas para manutenção e identificação de problemas nas vias é a inspeção manual, onde funcionários dos órgãos responsáveis procuram pelas malhas possíveis problemas, sendo esse um trabalho cujo o tempo de resposta para a correção do asfalto pode ser demorado, além de um alto custo e muitas vezes desperdiçado, uma vez que os funcionários possam, por sorte, focar somente em vias ainda em bom estado. Ainda que se identifique as vias com problemas asfálticos, a inspeção manual não permite classificar o grau de problemas e pode não vir a priorizar corretamente as malhas.

Existem diversos instrumentos que possibilitam a inspeção com dados mais precisos que viabilizam uma manutenção da malha asfáltica mais efetiva, isto é, identificam rapidamente os problemas. No entanto, tais equipamentos possuem um alto custo de aquisição e manuseio. Têm-se, por exemplo, o medidor de perfil portátil *Dipstick* [Dipstick 2017], um sensor preciso baseado na medição da inclinação, sendo ele operado por uma única pessoa. Para realizar uma medição com ele basta girar o aparelho cujo corpo principal é levemente menor que uma caixa de sapatos, sobre cada um de seus pés, apoiar novamente ambos os pés do aparelho no pavimento e aguardar cerca de 2 segundos entre cada giro seguindo uma trajetória definida. O custo deste equipamento é de mais de 11 mil dólares, além dos custos para seu manuseio, as medições abrange pequenas áreas e dependem de um técnico qualificado. Porém, existem perfilômetros inerciais (equipamento que mede o perfil da superfície mediante o contato) que permitem uma maior abrangência de medições, como o

nacional *CiberLaser* [Cibermétrica 2017], um veículo com sensores que captam a irregularidade da malha observada, sendo necessário um motorista habilitado para conduzi-lo. Ele permite ampla flexibilidade de configuração em relação ao número e ao tipo de módulos de medição. O módulo *CiberLogger* realiza a captura de dados em alta velocidade de módulos de medição laser, permitindo ao usuário o registro de observações em tempo real, vinculadas ao hodômetro. O módulo *CiberShell* é um *software* específico para cálculo dos índices de irregularidade longitudinal de pavimentos a partir de perfis capturados com o *software CiberLogger*. Por fim, o módulo *CiberLVC*, que é um programa específico para captura de imagens de ruas e rodovias a partir de um veículo em movimento para avaliação de quaisquer aspectos das imagens, seja do pavimento, seja da sinalização ou dos arredores. No entanto, esse módulo possui valores entre 250 e 400 mil reais.

Com a finalidade de aumentar a abrangência e a velocidade da coleta, bem como determinar os locais dos problemas de maior impacto na população, além de diminuir o uso de equipamento de alto custo e reduzir o esforço manual, foi proposto pelo Dr. Vinícius Mourão Alves de Souza o projeto “*Avaliação e monitoramento colaborativo das condições de ruas e estradas por meio de sensores de smartphones*”, o qual foi financiado pela Fundação de Amparo à Pesquisa do Estado de São Paulo (PIPE Nº 2016/07767-3). O objetivo desse projeto é desenvolver um aplicativo para *smartphones* que seja capaz de identificar automaticamente e em tempo real os problemas na malha asfáltica. Para isso, primeiro se coletam os dados dos acelerômetros dos *smartphones*. Posteriormente, essas informações serão submetidas a um classificador, o qual irá classificar a qualidade do asfalto com base nos dados do acelerômetro.

Uma forma viável de se gerar um classificador (ou modelo de classificação) é por meio de aprendizado de máquina [Mitchell 1997, Alpaydin 2004]. Um classificador é um modelo que prediz a classe de um objeto no qual não há essa informação. Essas classes são conhecidas também como rótulos ou categorias. O modelo de classificação é gerado a partir da extração de padrões das classes de um conjunto de dados, onde para treinar um classificador é necessário rotular previamente os dados de treinamento. No entanto, fazer a rotulação de dados brutos como os gerados pelo acelerômetro do celular pode ser uma tarefa difícil, tendo em vista que esses dados são apenas sequências numéricas que variam ao longo da série temporal, dificultando a identificação dos rótulos dessa sequência de valores. Sem auxílio de vídeos, imagens ou qualquer multimídia, a rotulação pode-se tornar inviável ou mesmo impossível de ser realizada para o problema em questão.

Para suprir essa necessidade, neste trabalho de conclusão de curso foi proposto um sistema que viabiliza o processo de rotulação dos dados oriundos do acelerômetro. Esse sistema consiste em sincronizar os dados numéricos gerados pelo acelerômetro com vídeos, além de exibi-los em um gráfico dinâmico para que o usuário possa selecionar uma sequência de dados para rotulação mais facilmente. A partir desse espaço de dados selecionado, é exibido o tempo do vídeo para então ser feita a rotulação. Foi proposto também uma tela de configuração, para que os rótulos fossem inseridos de acordo com a necessidade do usuário, podendo ele salvar e reutilizar rótulos criados por ele mesmo. Vale ressaltar que a ferramenta foi proposta para navegadores *web* com intuito de facilitar o uso para os usuários.

O restante deste trabalho está dividido da seguinte forma. Na Seção 2 serão apresentados os trabalhos relacionados, Na Seção 3 são apresentados os detalhes da *Ferramenta*

para rotulação de séries temporais geradas por dados de acelerômetros. Por fim, na Seção 4 serão apresentadas as considerações finais deste trabalho de conclusão de curso.

## 2. Trabalhos Relacionados

Há na literatura poucos trabalhos caracterizados pela seleção e rotulação de uma série temporal (ou segmento de uma série temporal). Dentre esses trabalhos pode-se citar [Team 2015], que aborda uma interface na linguagem de programação R para a biblioteca de gráficos em *JavaScript*. Ele fornece recursos para visualizar dados de séries temporais como zoom e realce na série temporal, barra inferior para selecionar um espaço na série temporal, além de permitir sobreposição no gráfico como a rotulação. Esta ferramenta possui uma funcionalidade em que o usuário pode clicar e arrastar em um determinado espaço temporal e rotulá-lo de acordo com sua necessidade.

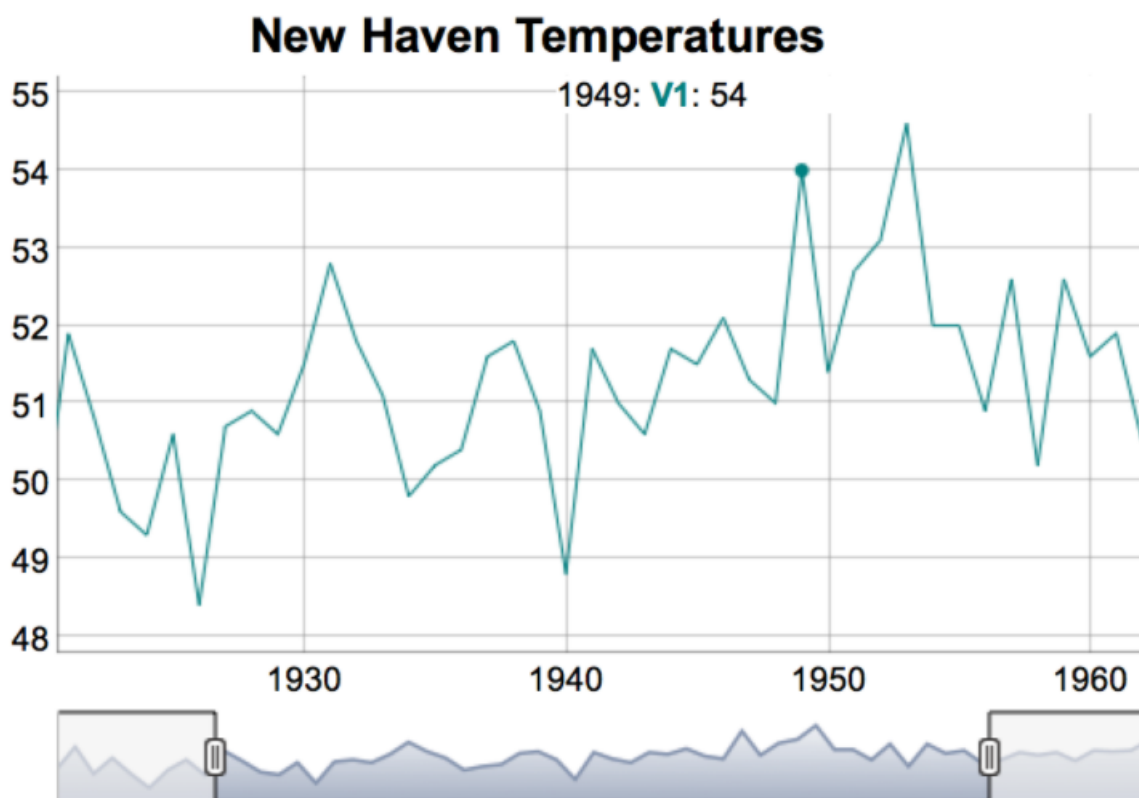
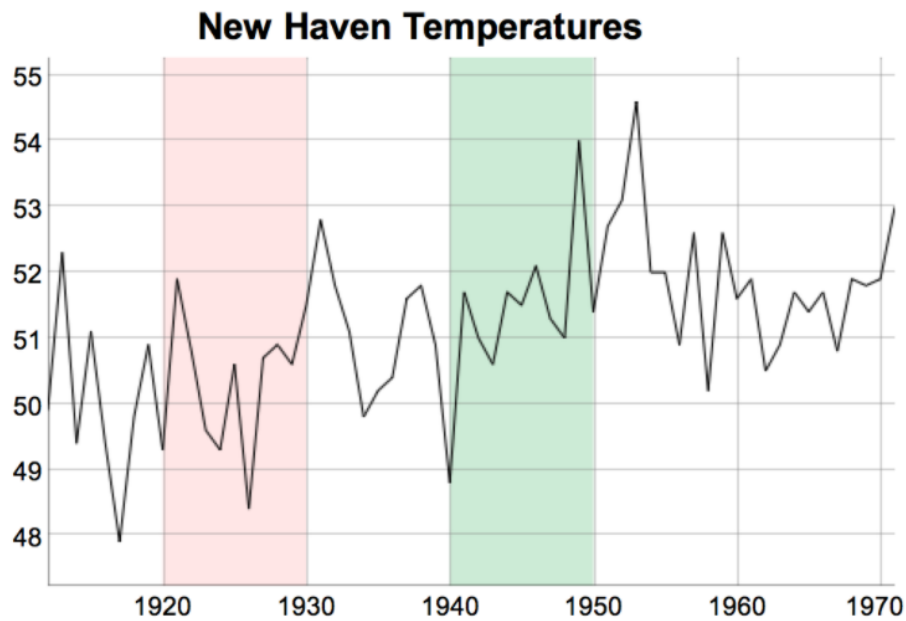
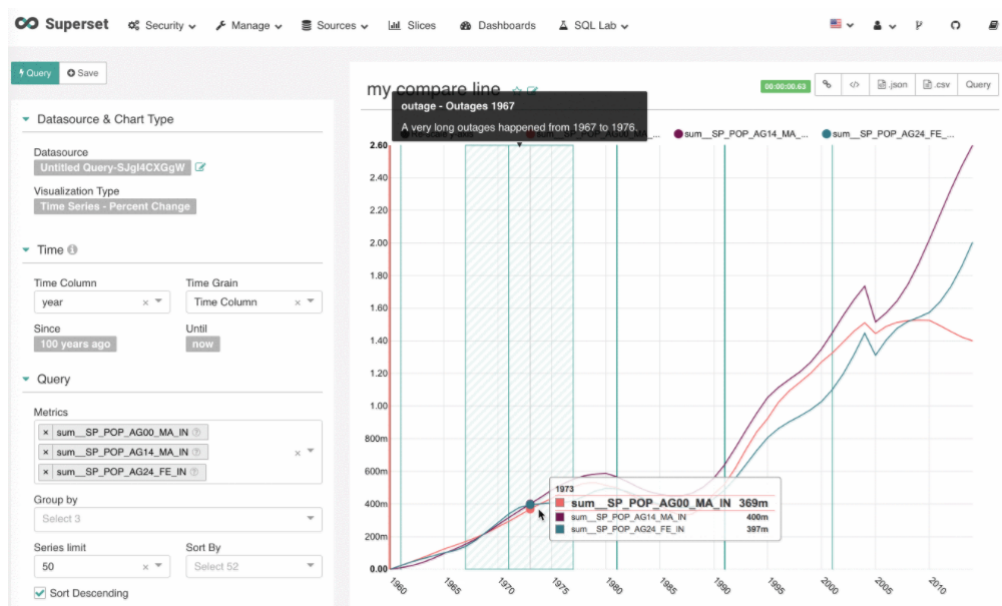


Figura 1. Gráfico da série temporal gerada pela ferramenta com opção de seleção de espaço [Team 2015].



**Figura 2. Momento da seleção de um espaço por sombreamento para rotulação [Team 2015].**

A *Web Application* [Caravel 2017] apresenta uma ferramenta de anotação de séries temporais por camadas em um gráfico. O usuário pode definir um ponto na série temporal ou um intervalo de tempo, rotulá-la e salvar como uma camada de anotação. Definida a camada de anotação, o usuário pode aplicar uma ou mais camadas no gráfico da série temporal para futuramente encontrar padrões semelhantes nos dados. A ferramenta ainda permite ao usuário verificar os dados rotulados nas camadas ao colocar o *mouse* sobre as camadas.



**Figura 3. Exemplo de camadas rotuladas em um gráfico de série temporal.**

Em [Röhlig et al. 2014] é proposta uma ferramenta de análise visual para apoiar o usuário na compreensão das influências dos parâmetros sobre os resultados gerados por algoritmos automatizados em séries temporais multivariadas. Algoritmos automatizados são ferramentas que computam os segmentos e atribuem rótulos para esses segmentos. O problema proposto neste artigo é a falta de uma configuração de apoio para escolha de parâmetros para um algoritmo de segmentação, que resultam em segmentos adequadamente rotulados. Com essa ferramenta, o usuário pode identificar e selecionar configurações de parâmetros que atendam a determinados critérios de qualidade. O *design* visual e interativo proposto ajuda a identificar relacionamentos e padrões temporais, suporta a tomada de decisões subsequentes e promove maior precisão e confiança na segmentação e rotulação.

Ao analisar os trabalhos relacionados, observa-se que estes não permitem a utilização de vídeos para auxiliar a rotulação. Isso torna difícil a rotulação de séries temporais oriundas dos acelerômetros, pois dificilmente o usuário conseguirá identificar, olhando somente a série, se um segmento da mesma corresponde à um buraco, lombada, ou trilho, por exemplo.

### 3. Proposta

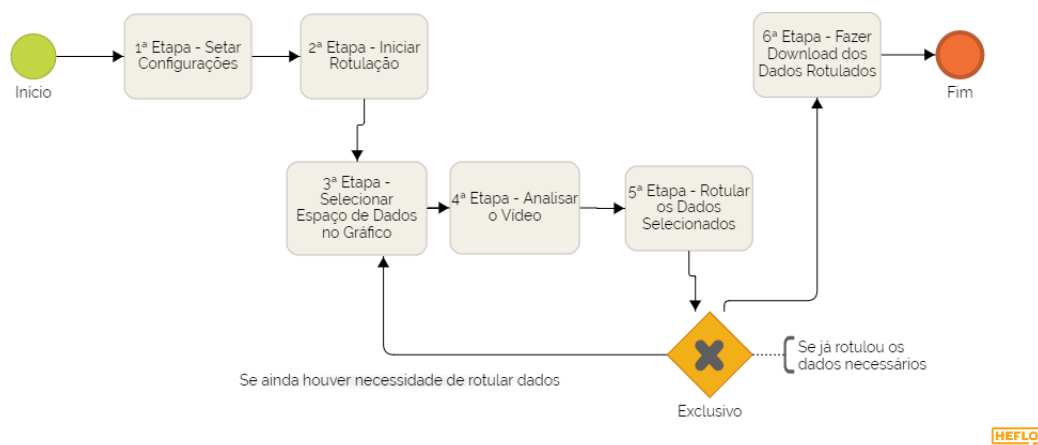
Nesta seção serão apresentadas a visão geral da ferramenta proposta, as técnicas e métodos utilizados e a descrição das etapas de funcionamento.

#### 3.1. Visão Geral

Foi proposto neste trabalho uma ferramenta *web* para rotular dados fornecidos por acelerômetros de *smartphones* no formato *Comma-separated values (CSV)*. Com intuito de tornar a rotulação mais fácil e rápida, a ferramenta auxilia o usuário por meio da sincronização de vídeos com os dados fornecidos.

Na Figura 4 são apresentadas as etapas do processo de rotulação de dados utilizando a ferramenta proposta. O processo de rotulação foi dividido em 6 etapas: (i) Setar Configurações, (ii) Iniciar Rotulação, (iii) Selecionar Espaço de Dados no Gráfico, (iv) Analisar o Vídeo, (v) Rotular os Dados Selecionados e (vi) Fazer Download dos Dados Rotulados. Os dados e vídeos devem ser previamente fornecidos pelo usuário por meio de uma tela de configuração. Nesta tela, também deverão ser fornecidos os rótulos, além da quantidade de quilômetros que serão exibidos por tela. Esses dados são exibidos em um gráfico dinâmico onde o usuário interage com ele e seleciona o espaço que será sincronizado com o vídeo correspondente, para então ser realizada a rotulação.

Para construção dessa ferramenta foram utilizadas as linguagens de programação *Javascript* e *PHP: Hypertext Preprocessor (PHP)* e a linguagem de marcação *HyperText Markup Language (HTML)*, além da utilização de um banco de dados relacional, o *MySQL*. Como servidor para testes foi utilizado o software *Xampp*.



**Figura 4. Esquema UML das etapas ao utilizar a ferramenta proposta.**

Nas próximas seções serão apresentados os detalhes de cada etapa do processo de rotulação implementado na ferramenta proposta.

### 3.2. 1ª Etapa - Setar Configurações

A primeira etapa do processo de rotulação de dados se inicia na tela de configuração, onde o usuário deve informar os possíveis rótulos, o caminho dos dados em um arquivo *Comma-separated values (CSV)* e o vídeo relativo aos dados. Um arquivo *CSV* é uma forma simples de armazenamento, que agrupa as informações de arquivos de texto onde cada coluna do arquivo *CSV* representa um atributo do conjunto de dados, cada linha representa uma instância do conjunto de dados e cada valor, separado por uma vírgula no arquivo *CSV*, representa uma mensuração de uma determinada instância para um determinado atributo. Por fim, se o usuário preferir dividir os dados em gráficos menores ele deve informar o valor de divisão. Para esta ferramenta, é necessário que o arquivo *CSV* possua os campos de tempo, a posição no sistema de posicionamento global, a magnitude, a latitude, a longitude, a velocidade, o deslocamento e um campo para inserir o rótulo. O campo magnitude, obrigatoriamente, deve estar na 5ª posição das colunas de atributos do arquivo. É possível nesta tela salvar a configuração para ser reutilizada posteriormente ou carregar uma configuração salva anteriormente.

Configurações Iniciais

INSIRA OS RÓTULOS PARA O DOMÍNIO

Digite aqui um rótulo

Adicionar

Salvar Configuração

Carregar Configuração

CARREGAR CSV:

Escolher arquivo Nenhum arquivo selecionado

CARREGAR VÍDEO:

Escolher arquivo Nenhum arquivo selecionado

Iniciar Rotulação

Figura 5. Tela de configuração inicial.

### 3.3. 2ª Etapa - Iniciar Rotulação

Na segunda etapa do processo de rotulação de dados é exibido ao usuário o vídeo, as opções de rótulo fornecidas por ele na etapa 3.2, e um gráfico com os dados em série temporal, como mostrado na Figura 6 e na Figura 7.

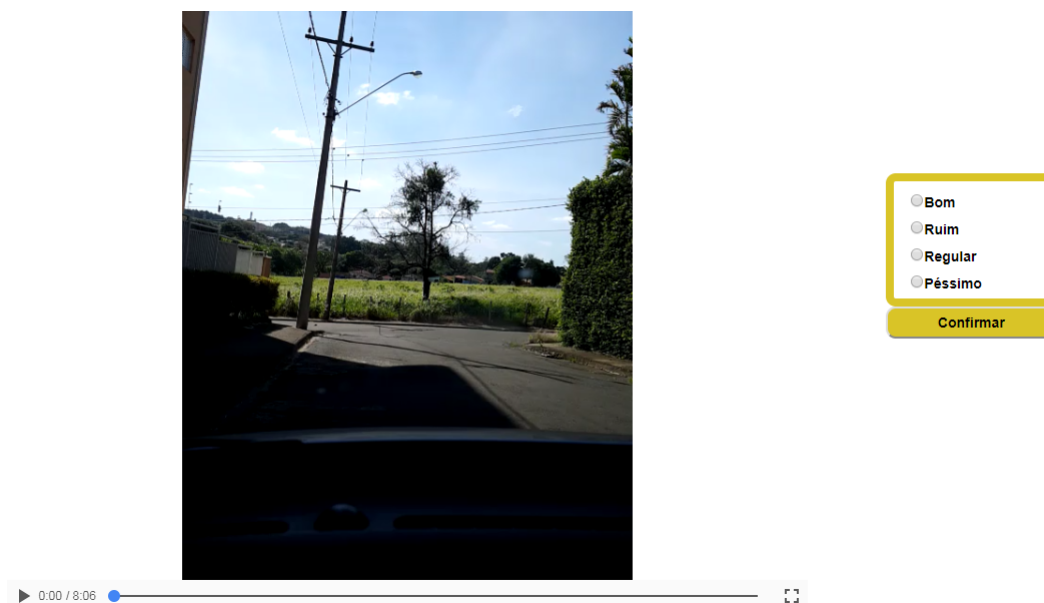
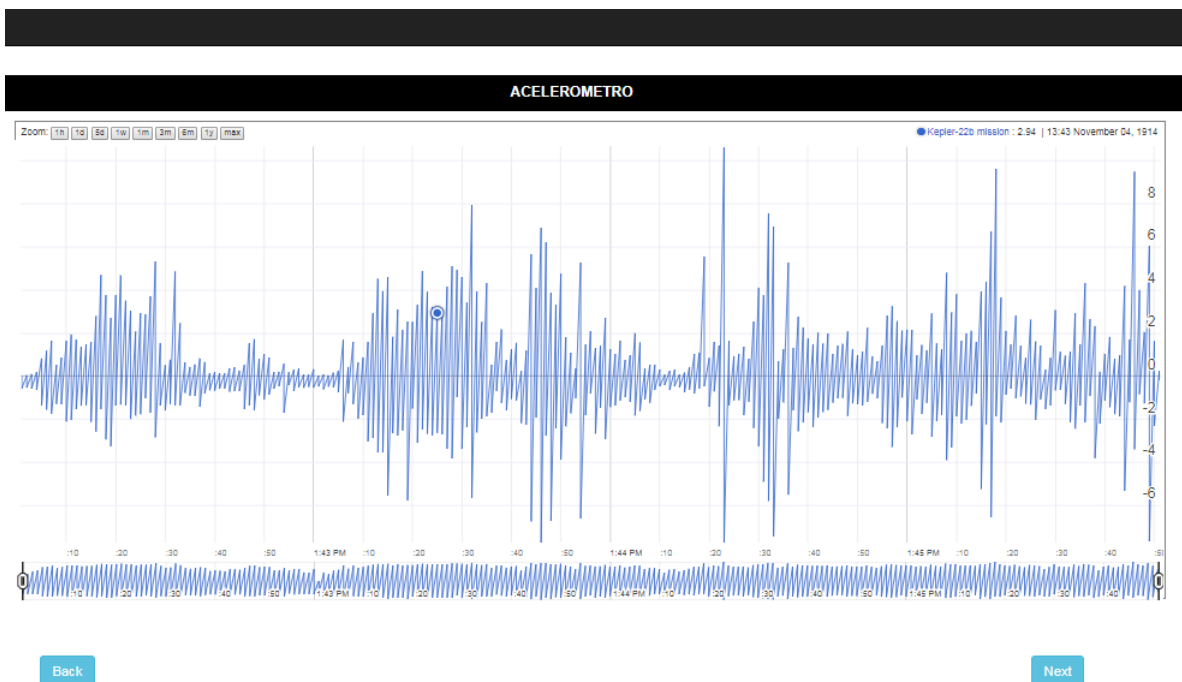


Figura 6. Imagem do vídeo limitado de acordo com o interesse do usuário.

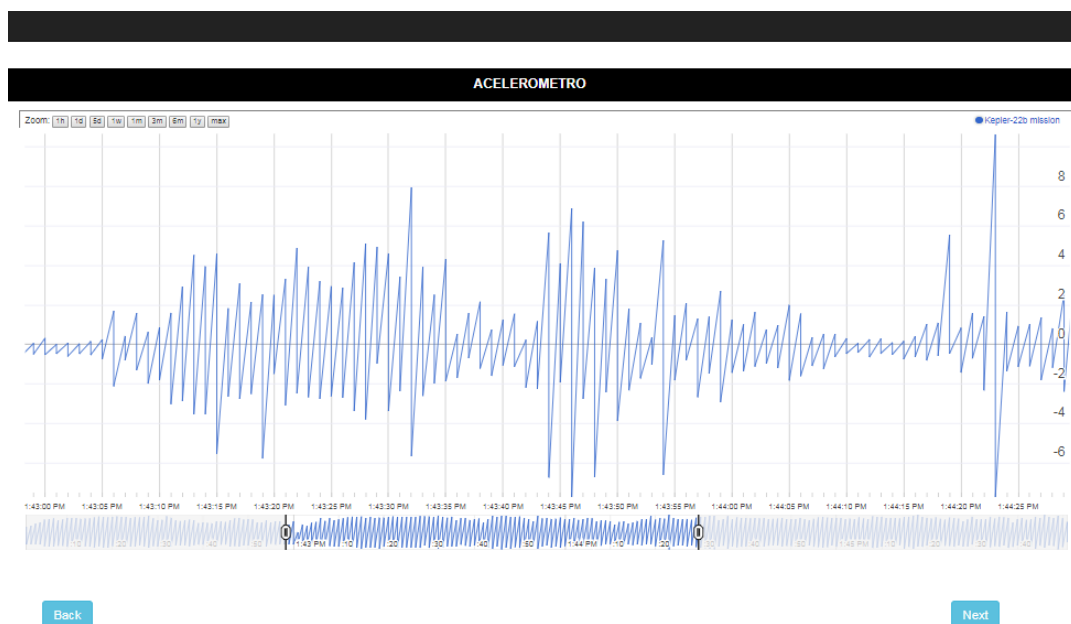




**Figura 7. Imagem do gráfico original sem interação.**

### 3.4. 3ª Etapa - Selecionar Espaço de Dados no Gráfico

Na terceira etapa do processo de rotulação de dados o usuário deve interagir com o gráfico e definir qual espaço será rotulado, como exibido na Figura 8. Ao interagir com o gráfico, automaticamente o vídeo é limitado ao trecho de dados selecionado, correspondendo o início e fim do vídeo ao início e fim do trecho selecionado.



**Figura 8. Imagem do gráfico após interação do usuário.**

### 3.5. 4ª Etapa - Analisar o Vídeo

Na quarta etapa o usuário deve consultar o vídeo que estará limitado de acordo com o espaço definido por ele na etapa 3.4, vide Figura 6. Foi proposto, nessa ferramenta, a utilização do *HTML5* para renderizar o vídeo.

### 3.6. 5ª etapa - Rotular os Dados Seleccionados

Após analisar o vídeo, o usuário rotulará o espaço de dados selecionado. Para realizar a rotulação, o usuário deve selecionar o rótulo dos dados na janela ao lado do vídeo, e confirmar a escolha no botão confirmar. Ao confirmar a escolha, a palavra "*Sucesso*" deverá ser exibida, como na Figura 9.

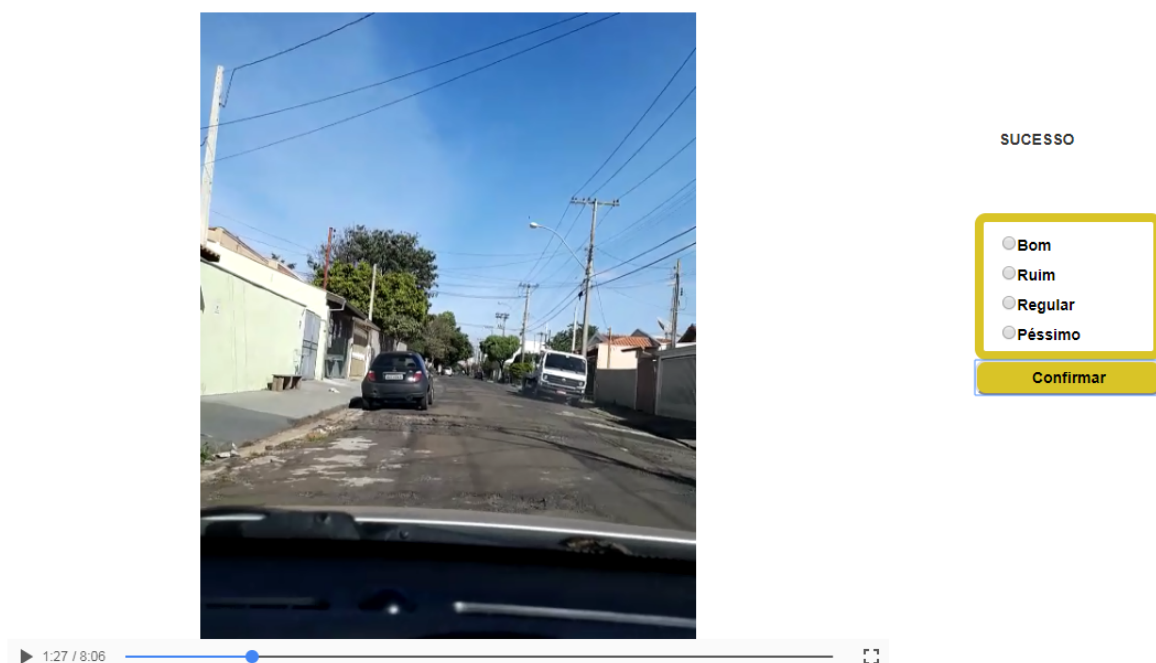


Figura 9. Imagem da ação de rotulação realizada pelo usuário.

### 3.7. 6ª Etapa - Fazer Download dos Dados Rotulados

A sexta e última etapa do processo de rotulação é obter a planilha rotulada com os rótulos determinados pelo usuário. O usuário deve iniciar o *download* clicando no botão "*Download da planilha rotulada*", vide Figura 10. Ao clicar é gerado um arquivo *CSV*, com as mesmas instâncias que foram utilizadas na entrada da ferramenta, porém, agora com um campo a mais que conterá o rótulo atribuído a instância, caso ela tenha sido rotulada.

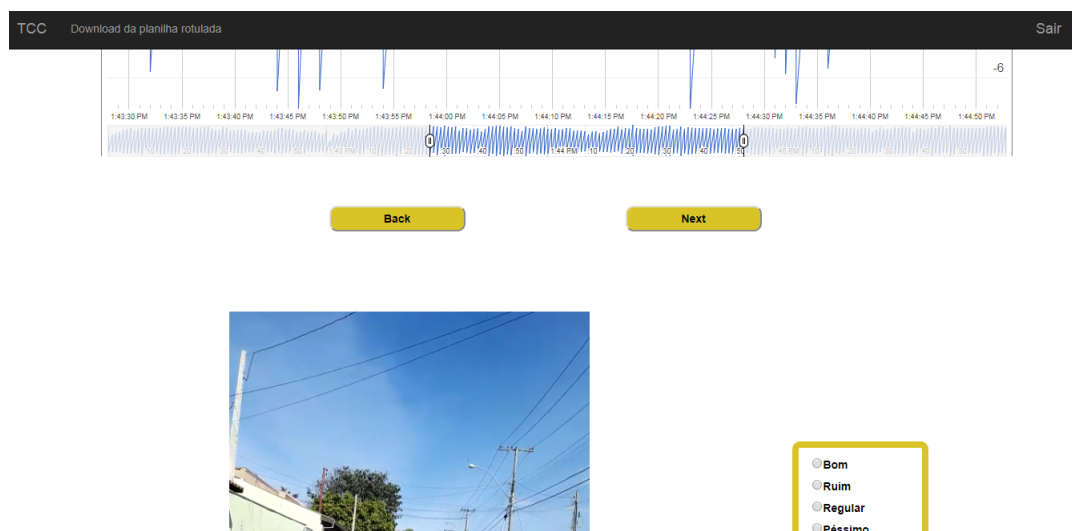


Figura 10. Imagem com o botão *Download da planilha rotulada* no menu superior.

#### 4. Considerações Finais e Trabalhos Futuros

Dada a proposta do projeto “*Avaliação e monitoramento colaborativo das condições de ruas e estradas por meio de sensores de smartphones*” surgiu a necessidade de uma ferramenta de rotulação a qual este trabalho de conclusão de curso supriu, trazendo uma ferramenta *web* onde, dada os arquivos *CSV* gerado pelo acelerômetro e o vídeo correspondente, o usuário pode rotular de forma mais rápida e precisa esses dados.

Como trabalhos futuros, pode-se trabalhar a evolução do aprendizado de máquina dentro da ferramenta, para detectar automaticamente os principais trechos da série temporal para o usuário rotular, como objetivo de diminuir o esforço de rotulação e consequentemente, aumentar a qualidade do modelo de classificação.

#### Referências

Alpaydin, E. (2004). *Introduction to Machine Learning*. Adaptive computation and machine learning. MIT Press.

Bazani, A. (2017). Frota circulante de veículos no brasil tem estabilidade e há menos ônibus nas ruas. <https://diariodotransporte.com.br/2017/05/19/frota-circulante-de-veiculos-no-brasil-tem-estabilidade-e-ha-menos-onibus-nas-ruas/>. Acessado em: 20 de novembro de 2017.

Brasil, P. (2014). IBGE mapeia a infraestrutura dos transportes no brasil. <http://www.brasil.gov.br/infraestrutura/2014/11/ibge-mapeia-a-infraestrutura-dos-transportes-no-brasil/>. Acessado em: 05 de novembro de 2017.

Brasil, P. (2017). Extensão de rodovias pavimentadas cresce 23% em 15 anos. <http://www.brasil.gov.br/infraestrutura/2016/05/extensao-de-rodovias-pavimentadas-cresce-23-em-15-anos/>. Acessado em: 05 de dezembro de 2017.

- Caravel (2017). Time series annotation layer · issue 3502. <https://github.com/apache/incubator-superset/issues/3502>. Acessado em: 05 de novembro de 2017.
- Cibermétrica (2017). Perfilômetro laser. <http://www.cibermetrica.com.br/index.html>. Acessado em: 14 de novembro de 2017.
- CNT, P. D. R. (2014a). Pesquisa CNT de rodovias 2015: Relatório gerencial. *Brasilia: CNT: SEST*.
- CNT, P. D. R. (2014b). Transporte desenvolvimento: Entraves logísticos ao escoamento de soja e milho. *Brasilia: CNT: SEST*.
- Dipstick (2017). Purchase a dipstick 2277 road profiler. [http://store.dipstick.com/store.php/dipstick/pd4162137/purchase\\_a\\_dipstick\\_2272\\_road\\_profiler](http://store.dipstick.com/store.php/dipstick/pd4162137/purchase_a_dipstick_2272_road_profiler). Acessado em: 14 de novembro de 2017.
- Mitchell, T. M. (1997). *Machine Learning*. McGraw-Hill, Inc., New York, NY, USA, 1ª edição.
- Röhlig, M., Luboschik, M., Schumann, H., Bögl, M., Alsallakh, B., and Miksch, S. (2014). Analyzing parameter influence on time-series segmentation and labeling. In *2014 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pages 269–270.
- Team, R. (2015). Interactive time series with dygraphs. <https://blog.rstudio.com/2015/04/14/interactive-time-series-with-dygraphs/>. Acessado em: 05 de novembro de 2017.